A phenomenon of opposing, rapid yet non-growing learning effects in verbal statistical learning

Lu Wang[a], Tianlin Wang[b], Wenbo, Yu[a], Dandan Liang[a,c]*

[a] School of Chinese Language and Culture, Nanjing Normal University, Jiangsu Province, P.R. China

[b] University at Albany, State University of New York, New York, USA

[c] Interdisciplinary Research Centre for Linguistic Science, University of Science and Technology of China, Anhui Province, P.R. China

* Corresponding author

Telephone: (+86) 13815866738

E-mail: ldd233@sina.com

Abstract

Learning effect in statistical learning (SL) is always refined as the ability to distinguish the target words from partwords in 2-alternative-forced task. However, this task did not answer how individuals represent target words and foils, thus may not be sufficient in providing independent learning effect on the items. Additionally, studies have rarely described the trajectory of each item's learning effect. The current study examined the independent learning effect of each type of words and discovered the pattern of learning trajectory in verbal SL task. Participants were randomly assigned to learn a continuous artificial speech stream in one of three conditions: a baseline, short-, or long-exposure time condition. Participants' learning was assessed using a familiarity ratings task. The different ratings between the baseline condition and the other two learning conditions were examined. Results revealed an opposing learning effect: familiarity ratings for target words were significantly higher than baseline, whereas foils' ratings were significantly lower. Additionally, there was no boost of learning effect for either the target or foil items as exposure time lengthened. This opposing, fast but non-growing learning effect not only suggest a complex mechanism underlying SL, but also provide insight regarding how to measure SL more efficiently.

Keywords: statistical learning, learning effect, learning trajectory, familiarity rating task

1. Introduction

Statistical learning (SL) encompasses the capacity to recognize statistical patterns and uncover cognitive units, constituting a fundamental aspect of cognition (Saffran et al., 1996). Within this context, verbal statistical learning (SL) has conventionally been acknowledged for its pivotal role in segmenting words within continuous speech. Ample research has linked SL and the development of language skills and reading abilities (e.g., Shoaib et al., 2018; Saffran & Kirkham, 2018; von Koss Torkildsen et al., 2019; Qi et al., 2019; Frost et al., 2020; Isbilen et al., 2022; Lukács et al., 2023). Yet, amid this progress, two questions persist: Firstly, what is the distinct learning effect of target words and foils in the realm of verbal SL? Secondly, how does this type of learning effect evolve as exposure time lengthens?

Independent learning effect of three types of words

The investigation of the independent learning effect of different word types within the context of statistical learning (SL) has been a focal point. Prior SL research has predominantly employed an offline learning paradigm, wherein participants listen to an artificial language during the exposure phase and subsequently engage in a 2-alternative-forced choice (2AFC) task. This approach has been applied to both children and adults (Wang & Saffran, 2014; Raviv & Arnon, 2018; Shoaib et al., 2018). In a typical verbal SL task, three distinct word types are involved:

1. Target words: These are the original nonsensical words used to construct the artificial language.

2. Partwords: Formed by combining consecutive syllables from two target words.

3. Nonwords: Created by utilizing non-adjacent syllables within the artificial language.

Multiple models have been put forth to explain the mechanics of SL, encompassing memory-based models (Thiessen et al., 2013; Thiessen & Erik, 2017; Endress, 2020) and chunking-based models (Isbilen et al., 2020; Isbilen et al., 2022). These models elucidate that SL triggers various processes tied to memory systems, including the activation, integration, and forgetting of information. Given that both target words and partwords embody statistical regularities and occur during the exposure phase, participants naturally form memory representations of these items.

Consequently, the observation that participants identify target words by contrasting learning performances between two alternatives highlights a confounding outcome. The scores from the 2AFC task represent a confounded result wherein the learning performance of both target words and foils potentially contributes. This interpretation underscores that the 2AFC task, while extensively used to detect SL learning effects, solely furnishes information about the ability to differentiate between target words and foils. It does not, however, elucidate how these individual words were processed or independently learned during the exposure phase. In essence, the 2AFC task, though a prevalent method, merely unveils relative learning effects between alternatives, lacking the differentiation of results based on item type independently (e.g., Mirman et al., 2008; Palmer & Mattys, 2016). In sum, there has been limited emphasis on examining the learning effects of each word type within the SL task. As a result, the manner in which individuals encode target words, partwords, and nonwords remains a domain yet to be explored.

Familiarity rating task and baseline condition

In measuring the effects of statistical learning (SL), the familiarity rating task emerges as a viable alternative to the conventional 2-alternative-forced choice (2AFC) task. This task has already been explored in prior studies, which have demonstrated a correlation between the

learning outcomes of this task and the 2AFC task. For instance, Batterink and Paller (2017) identified a linear decline in participants' learning across three distinct word types. Similarly, Erickson et al. (2016) established that the difference in rating scores between target words and partwords significantly exceeded zero. Notably, participants' performance in both the 2AFC task and the familiarity rating task exhibited correlation in specific versions of artificial languages.

Crucially, in contrast to the 2AFC task, the familiarity ratings task affords the opportunity to independently form memory representations for different word types. This facilitates participants in independently gauging their familiarity with each individual item. The present study adopts this approach to evaluate the independent learning effect on each word type.

Considering the aim of assessing the independent learning effect of the three word types, a direct comparison of familiarity rating scores between target words, partwords, and nonwords is unfeasible. The study also introduces a baseline condition, influenced by the work of Toro et al. (2011). In this baseline condition, the artificial language was composed of nonsensical syllables from the same pool used in the experimental condition. Since there were no instances of the three word types in the exposure phase of this condition, memory representations for these items were anticipated to remain at baseline levels. Consequently, the rating scores were deemed as the initial memory representation for each word type. The main contrasts in this study thus revolve around the rating differences for the three word types between the learning and baseline conditions.

Another rationale for incorporating a baseline condition arises from the need to address potential experiment-related effects within the artificial language learning paradigm. This concern is often addressed by employing two counterbalanced groups of participants, using different learning materials. By comparing the effects between these groups, researchers can

attribute the experiment's impact to the manipulated variables rather than a preference for arbitrary unit combinations. Similarly, in this study, the baseline condition serves to eliminate alternative explanations, with the absence of significant rating differences across the three word types suggesting that the design of the artificial language did not influence the experimental outcomes.

The trajectory of the learning effect

The temporal trajectory of the learning effect in statistical learning (SL) tasks has been a topic of investigation in previous studies. To ensure the detectability of the learning effect, many studies have employed extended exposure phases in their experimental designs. For instance, Toro et al. (2005) repeated each nonsensical word 150 times, and Wang and Saffran (2014) repeated each nonsensical word 130 times in a tonal artificial language. While some studies in the realm of event-related potentials (ERPs) have shown that effects like N100 or N400 occur later in the exposure phase, suggesting that longer exposure times are necessary for SL to take effect (Sanders, 2002; Abla et al., 2008; Batterink & Paller, 2017), the classic study by Saffran et al. (1996) demonstrated that infants could segment continuous speech and exhibit a significant learning effect after just a 2-minute exposure to an artificial language.

Contradictory findings have emerged regarding the timing of the SL learning effect. Recent studies have challenged the assumption that long exposure times are requisite for SL to occur. These studies have used relatively shorter exposure phases and still observed pronounced SL effects (Qi et al., 2019; Arnon, 2020). For instance, adults and children as young as 7 to 9 years old displayed SL effects after only 32 repetitions, and older children aged 8 to 16 years old showed SL effects after 48 repetitions.

Interestingly, a study by Siegleman and colleagues (2018) utilized a self-paced SL paradigm in the visual modality and shed light on the timing of SL effects. Their results revealed that in the visual domain, the learning effect followed a logarithmic function, and participants exhibited improved learning rates after as few as 7 repetitions of each triplet. However, caution is needed in directly applying this conclusion to verbal SL tasks, as some prior studies have pointed out differences in learning mechanisms between verbal and visual SL tasks (e.g., Frost et al., 2015; Frost et al., 2019; Emberson et al., 2019; Isbilen & Christiansen, 2022).

In summary, the timing of when the learning effect occurs during the exposure phase remains an open question, and few studies have explored the trajectory of the learning effect in verbal SL tasks. The existing research landscape presents varying viewpoints on whether the learning effect emerges early in the exposure phase or requires extended exposure times. Further investigation is needed to clarify this aspect of SL processes, particularly in the context of verbal SL tasks.

The current study

The primary goal of the present study was to delve into the independent learning patterns associated with three distinct types of words and to examine the trajectory of this learning process. To do so, the study manipulated three verbal statistical learning (SL) conditions:

1. Long-Exposure Learning Condition (LEL): In this condition, each nonsensical word was repeated a total of 90 times within the artificial language.

2. Short-Exposure Learning Condition (SEL): Within this condition, each word was repeated 45 times during the artificial language exposure.

3. Baseline Condition: The baseline condition served as a point of reference. It involved synthesizing syllables at random, without any occurrence of target words or partwords.

Within each of these SL conditions, three types of items were considered: target words, partwords, and nonwords. The study design employed a mixed-method approach, integrating both within-subject and between-subject variables. Specifically, SL condition (baseline condition, LEL condition, and SEL condition) functioned as the between-subject variable, while word type (target word, partword, and nonword) constituted the within-subject variable. To analyze the independent trajectory of the learning effect associated with target words and foils (nonwords and partwords), the study employed Linear Mixed Models (LMM) in the R statistical software. This analytical approach allowed for the exploration of how the learning effect evolves over time within each condition and for each type of word. By adopting this comprehensive experimental design and analytical framework, the study aimed to shed light on how different word types are autonomously learned and how this learning process develops across varying exposure times in the context of verbal SL tasks.

2 Method

2.1 Participants

One hundred and forty-four native speakers of Mandarin (age range:18-28; females = 123) were first recruited from a University in Southeast China. Participants were randomly assigned to either the long-exposure learning condition (49 participants), the short-exposure learning condition (49 participants) or the baseline condition (46 participants). All participants were right-handed with no formal musical training and were not majoring in foreign languages. The experiments were approved by the Institutional Review Board of the institution, and all participants signed informed consents before starting the experiments.

2.2 Materials

Twelve syllables were identified and combined with Tone 1 to create nonsensical tonal syllables following the approach in a study on Cantonese participants (Gómez et al., 2017). Syllables were recorded in a sound-attenuating room to digital format at 44100 Hz with 16bit precision. Target syllables were normalized for duration (350ms), mean pitch (266Hz), and intensity (70dB) via Praat software. Nonsensical words used in our artificial language were concatenated into syllable units.

The artificial languages in the two learning conditions were created with the same pool of target words. In the LEL condition, six target words were randomized to create an artificial language stream, which resulted in 90 tokens of each word. The same six words were used to create the artificial language in the LEL condition, which contained 45 tokens of each word with a fully randomized order of presentation. The LEL and SEL conditions were concatenated by a Praat script into a pseudorandom sequence, which ensured that the same word could not occur twice in a row. In the baseline condition, instead of consisting of six disyllabic words, the artificial language was concatenated with the same syllables that made up the other two conditions. The syllables were fully randomized in the baseline condition.

The within-word, syllabic level TP for target words was 1.0, and the TP of syllables spanning across word boundaries was 0.2. Nonwords consisted of two syllables that never co-occurred during the exposure. The within-word TP for nonword syllables was therefore always 0. The test items in the three conditions were identical, with a total of 18 items across three types of words. Three types of words are shown in Table 1. The artificial languages lasted about 6 minutes for the LEL condition and the baseline condition, but 3 minutes for the SEL condition. The data, materials and analysis code are available at https://osf.io/xh6ju/?view_only=6f1659f166934a47b4f5494aa4025dd1.

2.3 Procedure

All participants were told that they would hear an artificial language via headphones and would later be tested on their knowledge of the language. They then listened to the artificial language for either 6 or 3 minutes in a soundproof booth. After this exposure phase, a 6-point Likert scale familiarity rating task began (1 for not familiar at all and 6 for very familiar). Participants first took two practice trials, and then completed a total of 18 test trials. On each trial, participants were required to rate the familiarity of item considering the artificial language they had just listened to (see Fig.1). All three types of words occurred only one time, with the order of presentation randomized across trials.

3 Results

One nonword in the LEL condition was designed incorrectly for 11 participants and thus the data only included 17 trials for these participants, but 18 trials for others. To examine our main hypothese, a LMM (linear mixed model) were performed with function *lmer* in R[1]. The ANOVA results showed two significant main effects of condition ($F_{(2,140.87)} = 3.86$, $p = 0.02$) and word type ($F_{(2,15.02)} = 7.36$, $p < 0.01$) and a significant interaction effect ($F_{(4,2416.34)} = 26.39$, $p < 0.01$). The standardized Coefficients of Fixed effect could be seen in Table 2. We next ran a series of simple effect analysis with function *emmeans*. All p values were Bonferroni adjusted when pairwise comparisons consisted of more than two levels.

In order to rule out arbitrary preferences associated with artificial language; we first evaluated the comparability of the three types of test items in the baseline condition. Participants' rating scores of target words was not significantly different that of partwords and nonwords (target word – partword: $t = –0.70$, $\beta = –0.16$, $p > 0.05$); similarly, participants' rating scores of target words did not staticically differ from nonwords (target word – nonword: $t = –0.93$, $\beta =$ –

0.21, $p > 0.05$). Finally, there was no significant difference between partwords and nonwords' rating scores (partword – nonword: $t = –0.22$, $\beta = –0.05$, $p > 0.05$).

We then compared the rating difference between learning conditions and baseline condition to reveal items' independent learning effect and their trajectory. For target words, the rating scores in SEL condition ($M = 4.57$) was significantly higher than those of baseline condition ($M = 4.11$), $t = 2.82$, $\beta = 0.46$, $p = 0.02$. The rating scores in LEL condition ($M = 4.81$) was also significantly higher than those of baseline condition, $t = 4.25$, $\beta = 0.70$, $p < 0.01$. Then, target words' rating scores in LEL condition were not significantly different from those of SEL condition, $t = 1.45$, $\beta = 0.24$, $p = 0.44$.

Different from the above patterns, partcipants rated partwords more familiar in baseline condition ($M = 4.27$) than those in LEL condition ($M = 3.49$) and SEL condition ($M = 3.61$), baseline – LEL condition: $t = 4.74$, $\beta = 0.78$, $p < 0.01$, baseline – SEL condition: $t = 4.01$, $\beta = 0.66$, $p < 0.01$. The similar pattern was also found in the pairwise comparison of nonwords between three SL conditions (baseline: $M = 4.32$, LEL condition: $M = 3.53$, SEL condition: $M = 3.69$), baseline – LEL condition: $t = 4.82$, $\beta = 0.79$, $p < 0.01$, baseliene – SEL condition: $t = 3.79$, $\beta = 0.63$, $p < 0.01$. The rating difference of partwords ($t = 0.74$, $\beta = 0.12$, $p > 0.05$) and nonwords ($t = 1.02$, $\beta = 0.17$, $p > 0.05$) between LEL and SEL condition did not reach significance. See Fig. 2 for a visualization of participants' rating patterns in the three conditions. These results showed that participants have already achieved the learning effect at the beginning of exposure phase and this learning effect kept statble along the whole learning pahse.

4. Discussion

While previous research has established the occurrence of the learning effect in verbal statistical learning (SL) tasks among various participant groups, there has been a notable dearth

of studies examining the trajectory of independent learning effects across different types of words, including target words, partwords, and nonwords. Addressing this gap, the current study adopted a mixed-design experiment to delve into the nuanced learning dynamics present in verbal SL tasks. The findings of the study revealed an intriguing and contrasting pattern of learning effect. Specifically, participants demonstrated a swift yet non-progressive learning effect in the context of verbal SL tasks.

The study's results showed that participants exhibited higher levels of familiarity with target words while experiencing reduced familiarity with foils (both partwords and nonwords) during the short exposure condition in comparison to the baseline condition. Furthermore, this intriguing learning pattern was observed to be independent of the length of the exposure time. The lack of increase in learning effect with extended exposure time adds an additional layer of complexity to the nature of verbal SL processes. These findings contribute to our understanding of the intricate mechanisms underlying verbal SL and highlight the distinctive learning patterns associated with different types of words. The study's mixed-design approach, coupled with its exploration of the trajectory of learning effects, enhances the comprehensiveness of our insights into the dynamics of verbal SL.

4.1 The opposing learning trajectories of targets versus foils

Previous studies have commonly employed the 2-alternative-forced choice (2AFC) task, wherein participants decide between two options to determine familiarity or word presence in an artificial language, as a means to gauge statistical learning (SL) performance. While these studies have often shown participants' ability to differentiate between target words and partwords, they have rarely disentangled the distinct learning effects of different word types. The present study introduced a baseline condition to disentangle the independent learning effects of the three word

types. The findings revealed a notable pattern of familiarity ratings: target words exhibited significantly higher familiarity ratings in both short and long exposure conditions, while partwords and nonwords demonstrated substantial decreases in familiarity ratings from the baseline condition to the other two conditions.

This opposing pattern of familiarity ratings carries implications on two fronts. Firstly, it adds depth to our understanding of the components underlying the learning effects traditionally measured by the 2AFC task. Each trial's correctness in the 2AFC task can be further deconstructed into positive familiarity with target words and negative familiarity with partwords or nonwords. This suggests that the learning effects derived from the 2AFC task might overestimate the actual learning performance concerning target words.

Secondly, the positive learning effect exhibited by target words aligns well with both memory-based and chunking-based models of SL. These models propose that syllable units carrying greater statistical information are more likely to be recognized as target words and subsequently stored in memory. The relatively low transitional probability (0.2) between the words used in this study further supports this finding. Participants could easily segment speech into smaller chunks and rate them with higher familiarity compared to the baseline condition.

An important finding in the study was the markedly lower familiarity ratings for partwords and nonwords in the learning conditions relative to the baseline condition. The initial assumption was that partwords might receive slightly higher ratings due to their multiple repetitions during the exposure phase, while nonwords should exhibit consistent familiarity ratings across learning and baseline conditions due to their absence in the exposure phase. However, the contrary was observed. The study proposes that explicit mechanisms may contribute to these results. Previous research has indicated that supplementary explicit training enhances performance and elicits

distinct neural potentials (Batterink et al., 2015a). The role of domain-general resources like working memory in influencing verbal SL outcomes has also been highlighted (Palmer & Mattys, 2016). Given the task's relative simplicity for adults, it's conceivable that participants consciously memorized target words during the exposure phase, allowing them to explicitly reject foils during both exposure and test phases. This interpretation is in line with the notion of metacognition in the test phase of SL, as demonstrated by participants' higher metacognition levels in recognition trials involving target words over nonwords and phantom-word[2] over nonwords (Ordin & Polyanskaya, 2021). Collectively, these findings reveal a complex pattern of opposing learning effects, suggesting a confluence of both implicit and explicit mechanisms in SL processes (Batterink et al., 2015b).

4.2 The trajectory of the learning effect

Another key aim of the present study was to discern whether an extended exposure phase could augment the learning effect across the three types of words. Notably, the familiarity ratings for target words within the short exposure condition (SEL), wherein each word was repeated 45 times during the exposure phase, were already significantly higher than those in the baseline condition. Intriguingly, when the exposure time was doubled in the long exposure condition (LEL), with each word repeated 90 times, this did not yield a larger learning effect compared to the SEL condition. These findings highlight a distinctive pattern of learning characterized by rapid initial gains that do not significantly increase with prolonged exposure in the context of verbal SL.

The use of the familiarity rating task, as adopted in this study, contrasts with online SL tasks such as the target-detection task where participants identify the target syllable in real time as they learn the artificial language. Despite this difference, the observed fast-learning effect in the

current study aligns well with findings from other online studies. For instance, it has been demonstrated that following a single exposure to words within continuous nonsensical speech, participants exhibited faster reaction times (RTs) to final syllables compared to the initial syllables (Batterink, 2017). Notably, in a visual SL task, participants accelerated their learning pace after as few as 7 repetitions of triplets, indicating a rapid learning effect (Siegelman et al., 2018).

By contextualizing these outcomes, the present study contributes to the growing body of evidence supporting the notion of swift learning in both visual and verbal SL tasks. The congruence between these findings further underscores the intriguing nature of rapid learning effects within the realm of SL processes.

4.3 Limitations and future directions

The current study represents a preliminary effort in investigating the independent learning effects of target words and partwords, as well as elucidating the trajectory of these effects within the framework of a verbal statistical learning (SL) task. While the study successfully revealed an opposing pattern of learning effects and identified a rapid yet non-progressive learning trajectory, it still leaves a couple of important questions unanswered.

Firstly, the relationship between the learning effects derived from contrasting the learning condition with the baseline condition and the learning effects derived from contrasting target words with foils remains unexplored. In contrast to studies such as Batterink and Paller (2017), where the focus was on rating differences between target words and foils, the current study utilized the difference in ratings between learning and baseline conditions. Because the learning condition in this study was treated as a between-subject variable, conducting a direct correlation analysis between these two types of learning effects was not feasible. While a strong correlation

between them is anticipated, given their shared reflection of the process of tracking statistical regularities in speech, it would be valuable to investigate whether the learning effects recognized in prior studies can be deconstructed into distinct learning effects for different word types within the framework of this study. Secondly, the relationship between the learning effects of target words and partwords warrants further exploration. Notably, the comparison of learning effect differences revealed that partwords exhibited a slightly larger effect size than target words. This trend suggests that the independent learning effect of partwords might play a significant role in the SL performance traditionally revealed by the 2AFC task. The implications of this finding emphasize the need for additional research to delve into the potential interrelationships between various types of learning effects, shedding light on the complex mechanisms at play in SL processes.

Ultimately, the central finding of a rapid yet non-progressive learning effect in verbal SL tasks has far-reaching implications. It suggests the feasibility of employing concise artificial language structures to efficiently assess individuals' SL abilities. This observation holds particular significance in practical contexts, especially when investigating the links between SL ability and language development, especially in children. Recent literature has engaged in theoretical and psychometric discussions on this matter, and the current study's findings are poised to contribute to these ongoing discussions, offering practical insights that could prove valuable (as detailed in Siegelman et al., 2017).

5. Conclusions

The findings from this study provide valuable insights into the nature of the independent learning effect, which is influenced by the type of word being considered. Notably, the results demonstrate a distinct pattern: foils (partwords and nonwords) exhibit significantly lower

familiarity ratings, while target words exhibit notably higher familiarity ratings compared to the baseline condition. This divergence highlights the differential impact of word type on the learning effect.

Notes

1. model <- lmer(data = learning_effect, Rating ~ condition * word_type + (1|Subject) + (1|stimulus))

2. phantom-words: words which never occurred in the speech stream, but had exactly the same TPs as the target words that did occur in the speech stream.

Acknowledgments

This work was supported by the Social Science Foundation of Jiangsu Province Higher Education Institutions [2022SJYB2051]; and the Initial Scientific Research Fund of Nanjing Normal University [184080H202A121].

Declaration of Interest Statement

The authors report there are no competing interests to declare.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at [https://osf.io/xh6ju/?view_only=6f1659f166934a47b4f5494aa4025dd1.].

Reference

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952-964.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? an investigation of task reliability across modality. *Behavior Research Methods52* :68-81.

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological Science, 28*(7), 921-928.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31-45.

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015a). Functional differences between statistical learning with and without explicit training. *Learning & Memory*, 22(11), 544.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015b). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83*, 62-78.

Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2019). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism (s). *Developmental science, 22*(6), e12847.

Endress, A. D. , Slone, L. K. , & Johnson, S. P. (2020). Statistical learning and memory. *Cognition, 204*, 104346.

Erickson, L. C., Kaschak, M. P., ED Thiessen, & Berry, C. (2016). Individual differences in statistical learning: conceptual and measurement issues. *Collabra*, *2*(1), 14.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in cognitive sciences, 19*(3), 117-125.

Frost, R., Armstrong. B. C., & Christiansen. M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, 145(12), 1128-1153.

Gómez DM, Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: the role of lexical tones. *Language & Speech*, 61(1), 84-96.

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language: A Meta‐Analysis Into 25 Years of Research. *Cognitive Science, 46*(9), e13198.

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition, 225* (2022) 105123.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: a memory based approach to statistical learning. *Cognitive Science*, 44(7).

Lukács, Á., Dobó, D., Szőllősi, Á., Németh, K., & Lukics, K. S. (2023). Reading fluency and statistical learning across modalities and domains: online and offline measures. *Plos one, 18*(3), e0281788.

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, 108(1), 271-280.

Ordin, M., & Polyanskaya, L. (2021). The role of metacognition in recognition of the content of statistical learning. *Psychonomic Bulletin & Review, 28*, 333-340.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology, 69*(12), 2390-2401.

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading, 23*(1), 101-115.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's statistical learning abilities. *Development Science, 21*: e12593.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, 69(1), 181-203.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5*(7), 700-703.

Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? evidence from English‐learning infants hearing Italian. *Cognitive Science*, 42(8), 3083–3099.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: current pitfalls and possible solutions. *Behaviour Research Methods, 49*(2), 1-15.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cognitive Science, 42*(3), 692-727.

Thiessen, E. D, Kronstein, A. T, & Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792-814.

Thiessen, E. D., & Erik, D. (2017). What's statistical about learning? insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*.

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The Contribution of Language-Specific Knowledge in the Selection of Statistically-Coherent Word Candidates. *Journal of Memory and Language, 64*(2), 171–180.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-B34.

von Koss Torkildsen, J., Arciuli, J & Ona Bø Wie. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Difference*s, *69*(2019), 60-68.

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5.

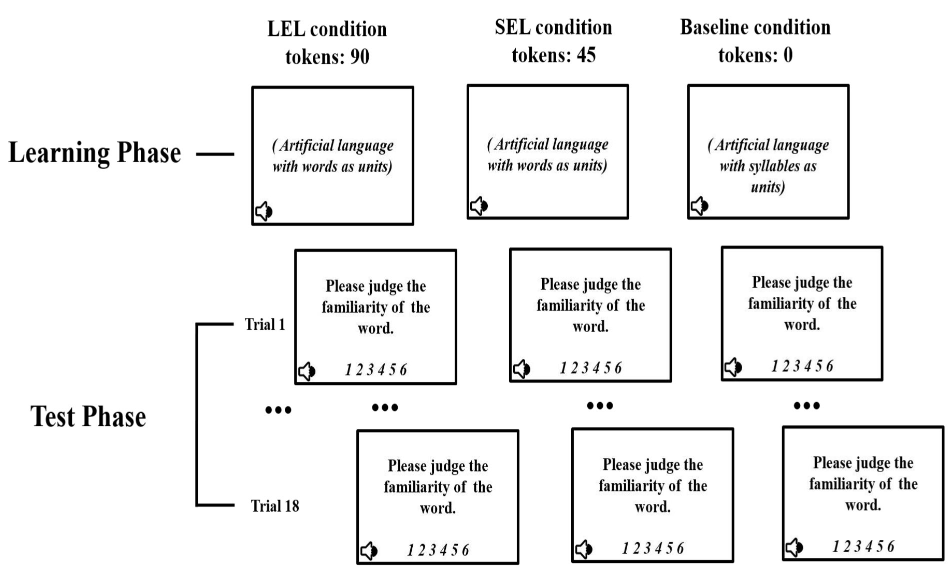
Table 1 test items in three SL conditions

target word	partword	nonword
meinei	semei	raore
raodia	diare	ruolai
ruose	neite	meite
laifo	nueruo	senei
tenue	forao	nuerou
rerou	roulai	fodia

Table 2. The standardized Coefficients of Fixed effect in LMM model (*estimate*, *SE*, *t* value, and *p* value)

Fixed effect	*Estimate*	*SE*	*t*	*p*
Intercept	4.11	0.18	22.79	< 0.001
conditionSEL condition	0.46	0.16	2.82	0.005
conditionLEL condition	0.70	0.16	4.25	< 0.001
word_type_partword	0.16	0.23	0.70	0.49
word_type_nonword	0.21	0.23	0.92	0.37
conSEL condition: word_type_partword	-1.24	0.17	-7.42	< 0.001
conLEL condition: word_type_partword	-1.36	0.17	-8.11	< 0.001

conSEL condition: word_type_nonword	-1.25	0.17	-7.50	< 0.001
conLEL condition: word_type_nonword	-1.32	0.18	-7.87	< 0.001



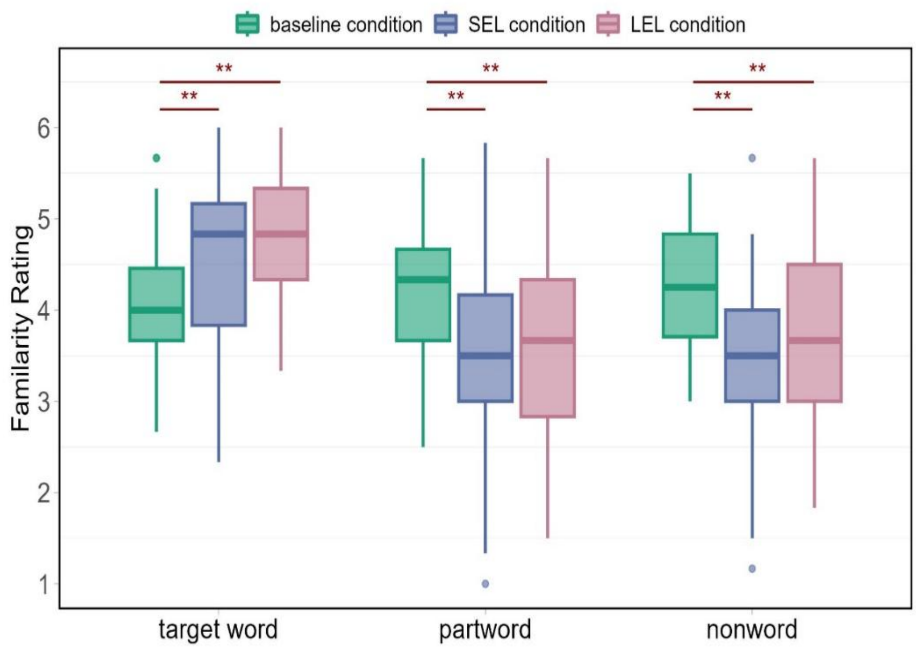


Figure	Figure Captions
Fig. 1.	Schematic representation of three conditions of verbal SL task
Fig. 2.	Familiarity ratings across word types in three conditions